Infection Control & Hospital Epidemiology (2019), **40**, 686–692
doi:10.1017/ice.2019.48

SHEA
The Society for Healthcare Epidemiology of America

Review

Design, implementation, and analysis considerations for cluster-randomized trials in infection control and hospital epidemiology: A systematic review

Lyndsay M. O'Hara PhD, MPH[1], Natalia Blanco PhD, MPH[1], Surbhi Leekha MBBS, MPH[1], Kristen A. Stafford PhD, MPH[1], Gerard P. Slobogean MD, MPH[2], Emilie Ludeman MSLIS[3] and Anthony D. Harris MD, MPH[1] for the CDC Prevention Epicenters Program

[1]Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, Maryland, [2]Department of Orthopaedics, University of Maryland School of Medicine, Baltimore, Maryland and [3]University of Maryland Health Sciences and Human Services Library, Baltimore, Maryland

Abstract

Background: In cluster-randomized trials (CRT), groups rather than individuals are randomized to interventions. The aim of this study was to present critical design, implementation, and analysis issues to consider when planning a CRT in the healthcare setting and to synthesize characteristics of published CRT in the field of healthcare epidemiology.

Methods: A systematic review was conducted to identify CRT with infection control outcomes.

Results: We identified the following 7 epidemiological principles: (1) identify design type and justify the use of CRT; (2) account for clustering when estimating sample size and report intraclass correlation coefficient (ICC)/coefficient of variation (CV); (3) obtain consent; (4) define level of inference; (5) consider matching and/or stratification; (6) minimize bias and/or contamination; and (7) account for clustering in the analysis. Among 44 included studies, the most common design was CRT with crossover (n = 15, 34%), followed by parallel CRT (n = 11, 25%) and stratified CRT (n = 7, 16%). Moreover, 22 studies (50%) offered justification for their use of CRT, and 20 studies (45%) demonstrated that they accounted for clustering at the design phase. Only 15 studies (34%) reported the ICC, CV, or design effect. Also, 15 studies (34%) obtained waivers of consent, and 7 (16%) sought consent at the cluster level. Only 17 studies (39%) matched or stratified at randomization, and 10 studies (23%) did not report efforts to mitigate bias and/or contamination. Finally, 29 studies (88%) accounted for clustering in their analyses.

Conclusions: We must continue to improve the design and reporting of CRT to better evaluate the effectiveness of infection control interventions in the healthcare setting.

(Received 29 November 2018; accepted 20 February 2019)

In a cluster-randomized trial (CRT), clusters or groups rather than individuals are randomized to interventions or treatments, and outcomes are measured in all (or a representative sample of) individuals in the clusters or groups. [1,2] CRT are well suited to evaluate public health, health policy, and health system interventions; they are ideal when the intervention carries a high risk of contamination. Contamination occurs when individuals randomized to different comparison groups are in close or frequent contact and may be influenced or "contaminated" by the intervention to which they were not randomized. This is likely to occur when comparing infection control and hospital epidemiology (ICHE) interventions within the same hospital or unit. Furthermore, when studying infectious diseases, individual randomization is often impractical because subjects in the nonintervention group may receive some protection due to the nature of transmission dynamics and herd immunity. Additional practical reasons for adopting this CRT design include simplified data collection, lower study costs, feasibility, ethical considerations, and often because the intervention is naturally applied at the cluster level.[1]

The CRT design has been well utilized in infectious disease research. Hayes et al[3] reviewed 21 papers that used a CRT design for infectious disease outcomes; however, all included studies described interventions applied solely in the community.[3] Wolkewitz et al[4] discussed a range of study designs, including CRT, that may be appropriate for intervention studies aiming to decrease hospital-acquired infections. Although the authors offer suggestions to improve the quality of such trials, the literature lacks specific examples of published studies that have used this approach in ICHE. The aim of this study is to present critical design, implementation, and analysis issues to consider when planning a CRT of interventions in the healthcare setting. Finally, we review and compare the reporting of CRT in ICHE to these established standards.

Author for correspondence: Anthony D. Harris, Email: aharris@som.umaryland.edu
Cite this article: O'Hara LM, *et al.* (2019). Design, implementation, and analysis considerations for cluster-randomized trials in infection control and hospital epidemiology: A systematic review. *Infection Control & Hospital Epidemiology*, 40: 686–692, https://doi.org/10.1017/ice.2019.48

© 2019 by The Society for Healthcare Epidemiology of America. All rights reserved.

https://doi.org/10.1017/ice.2019.48 Published online by Cambridge University Press

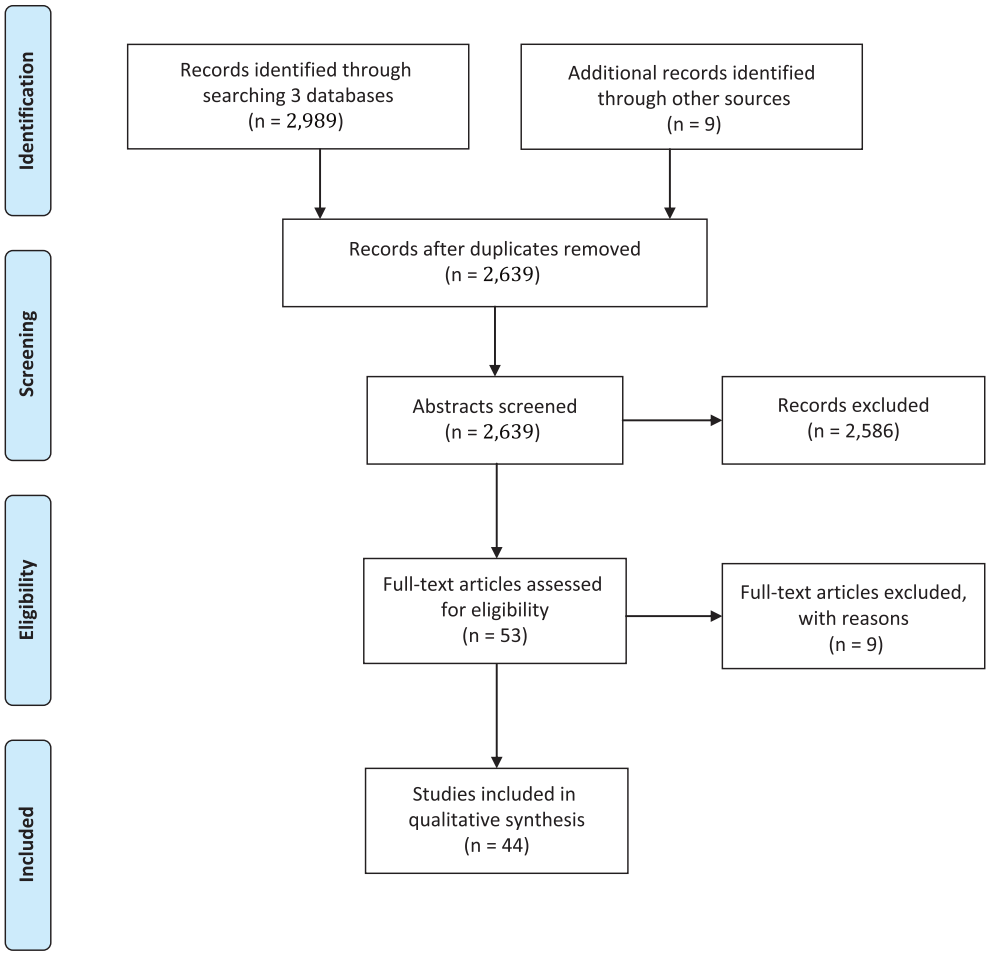


Fig. 1. PRISMA flow diagram* of search results. *From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000

Methods

Design, implementation, and analysis considerations

Identification of methodological principles. We identified 18 seminal review papers, expert papers, and textbooks on this topic published between 1981 and 2018. All authors reviewed these selected articles and relevant book chapters.[1–18] Each reviewer described their findings in 6 in-person group discussions. Lead author L.M.O. compiled recurrent themes. Finally, 7 epidemiological principles were deemed most important to CRT in the field of ICHE.

Systematic review of published cluster-randomized trials in ICHE

Search strategy

A search of 3 databases (Ovid MEDLINE, Embase, and CENTRAL) was conducted in June 2017 with a medical librarian (E.L.) to identify studies in the field of ICHE that utilized a CRT design. No date or language restrictions were utilized during the search process. An iterative process was used to generate the search terms and the general concepts and specific terms used (for details, see Appendix 1 online).

Assessment of studies

Full-text articles were reviewed independently by 2 investigators (L.M.O. and N.B.). To be eligible for inclusion, the study had to report an infection control outcome in the healthcare setting and had to employ a CRT design. Each study was assessed with respect to the 7 principles agreed upon. For each study, compliance with each methodological principle was recorded. Disagreements in compliance scoring were resolved by a third investigator (A.D.H.).

Results

After searching 3 databases, 2,989 records were identified and an additional 9 records were added by manually searching references of included articles. After removal of duplicates and elimination of articles based on title and abstract review, 53 full-text articles were reviewed. In total, 44 articles were deemed eligible for inclusion (Fig. 1). The most common reasons for exclusion were (1) the study setting was not healthcare; (2) randomization was not at the cluster level; (3) the primary outcome was not related to hospital infection prevention; and (4) the article did not present original research.

The 44 articles fell into the following topic categories: healthcare-associated infections (n = 18, 40.9%), antibiotic resistance (n = 10, 7%), hand hygiene (n=8, 18.2%), environment (n = 2, 4.5%), vaccination (n = 2, 4.5%), antibiotic stewardship (n = 2, 4.5%), and other (n = 2, 4.5%). The number of clusters enrolled ranged from 2 to 68 hospitals or units (for details and full summaries of the 44 included studies, see Appendix 2 online).

The following section briefly describes each epidemiologic principle and is followed by a description of the compliance of an included CRT to these principles.

https://doi.org/10.1017/ice.2019.48 Published online by Cambridge University Press

Table 1. Variations of the Cluster-Randomized Trial (CRT) Design

Type	Description	Advantages	Disadvantages
Parallel CRT	Clusters are randomly assigned to intervention or control in parallel	Straight-forward design and analysis	Clusters in the control arm do not have an opportunity to receive the intervention during the study period
Matched CRT	Matched pairs of clusters are randomly assigned to intervention or control. Examples of matching variables include hospital size, rural/urban location.	Addresses confounders at the design stage Offers some statistical advantage	If one cluster withdraws, the other hospital/ICU must also be dropped from the study.
Stratified CRT	Clusters are randomly assigned to intervention or control within a strata. Example: unit of randomization = infectious disease clinic, stratification variable = country or state	Ensures that an equal number of clusters from each group are in the intervention vs control groups Controls for confounding at the design stage	Requires a certain number of clusters in each group
CRT with stepped-wedge	All clusters eventually crossover but only from the control to the intervention at a time point determined at random.	Most appropriate when there is some evidence to support the intervention. All clusters eventually receive the intervention. Allows clusters to be enrolled gradually enrolled over time	Some argue that this is really a quasiexperimental design and is not truly randomized because no clusters get randomized to control.
CRT with crossover	All clusters receive both intervention and control in a sequence determined at random.	Reduces total number of cluster required	Requires a longer study duration Washout period is often not long enough resulting in bias.
CRT with crossover and multiple periods	All clusters receive both intervention and control multiple times in a sequence determined at random.	Reduces total number of cluster required Can assess carryover effect Allows determination of the true treatment effect vs a period effect	Adds another source of variance (between cluster, within cluster AND between cluster-periods) Requires longer study duration
Factorial CRT	Clusters are randomly allocated to 1 of 4 groups: receiving both interventions, intervention A, intervention B, or no intervention.	Can test whether intervention A is better than no intervention A or whether intervention B is better than no intervention B Because there is a combination of interventions A and B, it is also possible to test whether interventions A and B work together in synergy or are antagonistic. Used to assess the effects of 2 interventions in the same study and to explore interactions between interventions	Logistically and analytically complex
Fractional factorial CRT	Derived from a full factorial by dropping some conditions (combinations of factors) when resources are too limited to implement all possible combinations of factors or because some combinations cannot or should not be implemented	When conditions are removed, certain effects become completely confounded with each other and cannot be estimated separately.	Logistically and analytically complex

Principle 1: Design type and justification of use of CRT

The most basic form of CRT design is the parallel CRT; however, this design has several variations. Detailed descriptions, advantages, and disadvantages of each design type are outlined in Table 1. Authors should report the rationale for why the design chosen is most appropriate for their study. Some acceptable examples of design justification include the desire to minimize contamination bias between ICUs or floors in different study groups within the same facility, the recognition that a unit-level intervention would be more generalizable than randomly assigning the intervention at the patient-level, and the need to conduct a study of sufficient size with the available resources.

Systematic review findings

Of the 44 studies included in the review, 15 (34.1%) used a CRT with crossover, 11 (25.0%) used a parallel CRT design, 7 (15.9%) used a stratified CRT design, 4 (9.1%) used a CRT with stepped-wedge design, 3 (6.8%) used a matched CRT design, 2 (4.5%) used a CRT with crossover and multiple periods, and 2 (4.5%) used a stratified CRT design with crossover. Also, 22 of the included studies (50.0%) offered justification for their use of a CRT (Table 2). In a good example of justification for the CRT design, Huang et al (2016) stated that they " . . . chose this design to obtain results that could be generalized to the broadest set of hospitals, to use processes potentially adoptable by many hospitals, and to conduct a study of sufficient size with the available resources. Randomization of entire hospitals allowed us to recruit a broad array of hospitals" (see Appendix 2 online).

Principle 2. Accounting for clustering when estimating sample size and reporting of intracluster correlation coefficient or coefficient of variation

The correlation and thus nonindependence that exists among individual patients in a cluster must be accounted for when estimating sample size for such trials, yet many studies neglect to consider the within-cluster and between-cluster variation as measured by the intracluster correlation coefficient (ICC) or coefficient of variation (CV). A review by Simpson et al[5] of primary prevention trials

https://doi.org/10.1017/ice.2019.48 Published online by Cambridge University Press

Table 2. Compliance with 7 Key Epidemiological Principles Among Published Cluster-Randomized Trials (CRT) in Infection Control and Hospital Epidemiology (N=44)

Epidemiological Principle	Characteristics	No. (%)
1	**Type of design used**	
	CRT with crossover	15 (34.1)
	Parallel CRT	11 (25.0)
	Stratified CRT	7 (15.9)
	CRT with stepped wedge	4 (9.1)
	Matched CRT	3 (6.8)
	CRT with crossover and multiple periods	2 (4.5)
	Stratified CRT with crossover	2 (4.5)
	Factorial CRT	0
	Fractional factorial CRT design	0
	Justified use of CRT	22 (50.0)
2	**Sample size estimates**	
	Accounted for clustering when estimating sample size	20/33 (60.6)
	Reported design effect (ICC or CV)	15/33 (45.5)
3	**Consent**	
	Obtained waived consent	15 (34.1)
	Did not report how they dealt with consent	14 (31.8)
	Obtained consent from individuals	8 (18.2)
	Obtained consent at the cluster level	7 (15.9)
4	**Level of inference**	
	Individual level	32 (72.7)
	Cluster level	11 (25.0)
	Both individual and cluster level	1 (2.3)
5	**Matching and/or stratification**	
	Employed matching or stratification at time of randomization	17 (38.6)
6	**Bias and/or contamination**	
	Reported some effort to reduce bias/contamination	34 (77.3)
7	**Analysis**	
	Accounted for clustering in the analysis	29/33 (87.9)

showed that only 4 studies (19%) accounted for between-cluster variation in their sample size or power calculation. ICC measures the degree of similarity among outcomes within a cluster.[6] Generally, the higher the ICC, the more similarity that exists within clusters resulting in a loss of precision estimating effect of intervention. Therefore, standard approaches for estimating sample size that do not consider clustering may increase the probability of a type II error, meaning that the study will be underpowered.[1]

In some studies, clustering may arise at >1 level; therefore, 2 ICCs should be defined, for example, when an ICU within a hospital and the hospital itself are randomized. Variation exists among hospitals in addition to variation among ICUs within a hospital. An additional source of variance arises when the crossover design is used and each cluster receives the intervention in a separate period of time. In this case, it is important to account for period variance.[7]

Cluster randomization is less statistically efficient than randomizing individuals. Increasing the number of clusters enrolled in a CRT has a greater impact than increasing the number of individuals enrolled within each cluster on statistical power.[6,8] Therefore, many investigators choose to enroll a subsample of individuals within each cluster. The numbers of individuals needed to enroll per cluster depends largely on the underlying value of the ICC[9] and the anticipated effect size. A paper by Rutterford et al[10] provides detailed guidance on how to estimate sample sizes for CRT.

Systematic review findings

As shown in Table 2, 20 of 33 studies (60.6%) in which inference was made at the individual level accounted for clustering at the design phase when estimating sample size and power for their study. In addition, 15 studies (45.5%) reported the ICC, CV, or design effect. These values ranged from 0.005 to 0.38.

Principle 3. Consent

Randomization of groups rather than individuals presents unique ethical considerations. It may be appropriate for key decisions

https://doi.org/10.1017/ice.2019.48 Published online by Cambridge University Press

Table 3. Summary of Key Design and Analysis Considerations When Developing a Cluster-Randomized Trial (CRT) in Infection Control and Hospital Epidemiology

Epidemiological Principle
1- Design • Select the appropriate type of CRT (see Table 1) • Report the justification/rationale for using this design in the introduction section of the paper
2- Sample size estimates • Account for clustering by including a design effect such as ICC or CV in the sample size estimate • Report the ICC or CV used to estimate sample size • Report the effect size used to estimate sample size
3- Consent • Consider seeking consent at the cluster level • Determine whether it is appropriate and feasible to seek consent from representatives of the cluster rather than from individuals within each cluster
4- Level of inference • Define whether the level of inference is at the individual or the cluster level
5- Matching and/or stratification • Consider using an appropriate matching factor to improve power • However, use caution when employing this approach, especially if the number of clusters is large • Consider using a stratified design if appropriate
6- Bias and/or contamination • Consider the following techniques to reduce the potential for bias and/or contamination: Use an appropriate wash-out period if using a crossover design Implement the study in areas where clusters are distinct and well separated Use control group clusters that are external to the experimental trial Use multiple crossover periods
7- Analysis • If the level of inference is the individual, account for clustering in the analysis • Consider using statistical techniques such as mixed-effects models

makers to act as surrogates for a community or cluster and consent to randomization.[11] For example, nurse managers may consent on behalf of their unit to participate in an intervention trial with the outcome of hand hygiene adherence. Although ethical approval may be given at the cluster level, the refusal of an individual patient or healthcare worker (HCW) to participate in a study must be Epidemiology respected. It can be logistically difficult and perhaps unfeasible to obtain individual consent from large clusters.[12]

Systematic review findings

Overall, 15 studies (34%) obtained waived consent, 14 (32%) did not report how they dealt with consent, 8 (18%) reported that they obtained consent from individuals, and 7 (16%) reported consent at the cluster level. A good example of consent at the cluster level is described by Fuller et al (2012) where ward managers, infection control nurses, and ward coordinators consented on behalf of all other staff members to participate in a hand hygiene study (see Appendix 2 online).

Principle 4. Level of inference

In epidemiology, inference refers to the statistical process of generalizing from sample data to a wider population. A key property of CRT is that inferences are frequently intended to apply at the individual level, whereas randomization occurs at the cluster or group level. For example, to evaluate the effectiveness of a hand hygiene improvement intervention, researchers may choose randomization to occur at the unit level but adherence with hand hygiene recommendations to be assessed for each individual HCW within each cluster.

It is important to correctly identify whether the unit of inference will be at the individual or cluster level early in the planning stage of the trial. If randomization, variable collection, and analysis are all conducted at the cluster level, then sample size estimates and statistical analyses can be done as a standard randomized controlled trial.[6]

Systematic review findings

Of the 44 included studies, the level of inference was considered at the individual level for 32 studies (72.7%) and at the cluster level for 11 studies (25.0%). In 1 study (2.3%), randomization, variable collection, and analysis were conducted at both the individual and cluster levels.

Principle 5. Matching and/or stratification

Although matching can provide a simple method to consider potential confounders at the design stage, this approach may be overused and effective matching may be especially difficult in smaller studies.[13] Recruiting a large number of pairs provides statistical advantage only if the pairs represent different levels of baseline risk.[1] Furthermore, if a single member of a matched pair drops out of the study, this requires that both members of the pair be dropped from the analyses, thereby possibly rendering the study underpowered. Matching in a CRT should therefore be adopted with caution. Stratification is another approach that is commonly used to ensure that there is balance in cluster size per intervention and control groups within strata.[1,6]

Systematic review findings

Overall, 17 studies (38.6%) matched or stratified at time of randomization, whereas 27 (61.4%) did not employ either of these techniques. Examples of matching variables used included geographic region, rate of outcome, type of ICU, number of ICU or hospital beds, and hospital volume. A good example of appropriate matching can be found in the BUGG study published by Harris

https://doi.org/10.1017/ice.2019.48 Published online by Cambridge University Press

et al (2013), in which ICUs were paired and matched based on baseline MRSA and VRE acquisition rates (see Appendix 2 online).

Principle 6. Reducing the potential for bias and/or contamination

The goal of randomization is to minimize bias or to ensure that the baseline characteristics of the various clusters are balanced in different intervention groups. When conducting a study in the healthcare setting, the "transmission" of behaviors, attitudes, or knowledge among HCWs who are in regular contact can result in similar responses. This is sometimes referred to as a "herd effect." Similarly, the Hawthorne effect can be an issue in CRT. Intervention groups may benefit from increased attention and not solely from the intervention itself. To mitigate this, instead of studying only the standard of care in the control group, researchers may consider using a "minimal intervention" or "active controls." Puffer et al[14] found potential recruitment bias in 14 of 36 CRT reviewed (14%). There are several additional ways to reduce the potential for bias and contamination when conducting a CRT in the field of ICHE. For example, the study can be implemented in areas where clusters are distinct and well separated, and control-group clusters can be used that are external to the experimental trial; randomizing different locations within a hospital to control and intervention groups may be problematic. If a crossover design is used, it may be appropriate to employ multiple crossover periods and a wash-out period that is long enough to ensure that there are no residual effects. Furthermore, employing the CRT with crossover design is only appropriate if there is no carryover, which is rare in ICHE.

Systematic review findings

Overall, 34 studies (77.3%) reported some efforts to reduce the potential for bias and/or contamination. Most common were the use of a baseline period and the use of a wash-out period. Of all 44 studies, 7 (16.5%) reported the use of a baseline period, and 7 of 19 studies that used a crossover design (36.8%) reported using a wash-out period, which ranged from 2 to 4 weeks. Also, 3 studies (6.8%) specifically reported that the intervention was implemented in clusters that were distinct and well separated. A good example of efforts to reduce bias and contamination are described by de Smet et al (2009) (see Appendix 2 online). The authors ensured that the order of digestive tract decontamination regimens were randomly assigned, that the person in charge of randomization was blinded to ICU identity, and that the study periods were preceded by a wash-in and/or wash-out month.

Principle 7. Accounting for clustering in the analysis

The lack of independence among individual patients or HCWs in the same cluster, creates special methodological challenges. If between-cluster variation is not taken into account, a false claim of statistical significance may result via an increase in the probability of a type I error. Therefore, a main concern in CRT is internal validity. Many CRT fail to account for between-cluster variation at both the design and analysis stage. The aforementioned review by Simpson et al of primary prevention trials showed that only 12 (57%) accounted for clustering in their analyses.[5]

To obtain unbiased estimates of the effect of the intervention, analyses must be based on data from all cluster members or must be based on a random subsample of cluster members. It is necessary to decide whether to model the predictor variables as either fixed or random.[15] In many CRT, the cluster effect is modeled as random and the intervention effect is modeled as fixed. Several different approaches can be used to ensure that all comparative analyses allow for the clustered nature of the data and that correct confidence intervals and type I error rates are calculated. For example, a generalized estimating equation (GEE) can accommodate cluster-level and individual-level covariates. Similarly, proportional-hazards models with shared frailties can account for clustering within hospitals.

Systematic review findings

Of 33 studies, 29 (87.9%) accounted for clustering in their analyses. Only those in which the level of inference was the individual were included in the denominator of this calculation. Most of these studies used mixed-effects regression models or fixed-effects regression models to account for clustering. Rupp et al (2008) explain how they accounted for clustering at the analysis stage as follows: "... GEE were used to analyze hand hygiene adherence rates over time and their relationship to job category and hand gel availability, appropriately accounting for the potential correlation among observations" (Appendix 2 online).

Discussion

We have presented 7 critical design, implementation, and analysis principles to consider when planning a CRT of infection prevention and control interventions in the healthcare setting (summarized in Table 3). Adherence to these principles was variable among 44 ICHE studies identified by a systematic review, which suggests the need for more systematic reporting in this field. Notably, we did not identify any published studies in this field that employed a factorial or fractional factorial design. As shown in Table 1, each design type has advantages and disadvantages. The most appropriate design depends on the setting and research question. Many studies (82%) reported accounting for clustering during their analyses; however, <50% reported accounting for clustering when estimating sample size, and only 34% reported the ICC or CV that they used to do so. Reporting of these design effects is necessary to provide references for what constitutes a reasonable estimate for similar interventions and outcomes.

The aforementioned review, conducted in 2000, assessed CRT of infectious disease outcomes and identified only 21 such studies.[3] Our study included twice as many published articles, even when narrowed to a small subset of infectious diseases research. This illustrates the emergence of this design in research in recent years. Another recent review examined CRT in the general practice setting that included a patient-relevant outcome.[16] This article suggests that when studies of complex interventions (like those in the healthcare setting) are poorly designed and implemented, they often do not yield useful information. Because CRT are complex and costly, methodological rigor is of utmost importance.

In addition to the epidemiological principles presented here within the context of ICHE research, several tools are available to improve CRT. The Consolidated Standards of Reporting Trials (CONSORT) checklist provides evidence-based recommendations for reporting randomized trials and encourages authors to report their work in a transparent and standardized manner. CONSORT now offers an official extension for CRT,[17] and researchers are encouraged to refer to this. Similarly, Hemming et al[18] present power and precision curves that can be used as guidance when determining cluster size and Reich et al[19] provide a framework and R code for estimating power via simulation with or without 1 or more crossover periods. Finally, Caille et al[20] developed a graphical tool that

https://doi.org/10.1017/ice.2019.48 Published online by Cambridge University Press

identifies potential bias in CRT by depicting the time sequence of steps and blinding status.

In conclusion, the CRT design is used often in the field of ICHE, yet adherence to critical epidemiological principles remains suboptimal. Conduct and reporting of methodologically rigorous evaluations of infection prevention and control outcomes in the healthcare setting can inform best practice and policy.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/ice.2019.48

Author ORCIDs. Natalia Blanco, 0000-0002-3157-1119

Financial support. This research was funded by the National Institutes of Health (grant no. K24AI079040-05 to Dr Harris), by the CDC Prevention Epicenter Program (grant no. 1U54CK000450-01) and by the Banting Postdoctoral Fellowship Program administered by the Government of Canada (to Dr O'Hara).

Conflicts of interest. All authors report no conflicts of interest relevant to this article.

References

1. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London: Arnold, 2000.
2. Moberg J, Kramer M. A brief history of the cluster randomized trial design. JLL Bulletin: Commentaries on the history of treatment evaluation. James Lind Library website. http://www.jameslindlibrary.org/articles/a-brief-history-of-the-cluster-randomized-trial-design/. Published 2015. Accessed March 8, 2019.
3. Hayes RJ, Alexander ND, Bennett S, Cousens SN. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Statist Method Med Res* 2000;9:95–116.
4. Wolkewitz M, Barnett AG, Martinez MP, Frank U, Schumacher M, IMPLEMENT Study Group. Interventions to control nosocomial infections: study designs and statistical issues. *J Hosp Infect* 2014;86:77–82.
5. Simpson JM, Klar N, Donnor A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. *Am J Pub Health* 1995;85:1378–1383.
6. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Pub Health* 2004;94:416–422.
7. Arnup SJ, McKenzie JE, Hemming K, Pilcher D, Forbes AB. Understanding the cluster randomized crossover design: a graphical illustration of the components of variation and a sample size tutorial. *Trials* 2007;18:381.
8. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*, Third Edition. Philadelphia: Lippincott, Williams & Wilkins; 2008.
9. Donner A, Birkett N, Buck C. Randomization by cluster sample size requirements and analysis. *Am J Epidemiol* 1981;114:906–914.
10. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. *Int J Epidemiol* 2015;44:1051–1067.
11. Sim J, Dawson A. Informed consent and cluster-randomized trials. *Am J Pub Health* 2012;102:480–485.
12. Edwards SJL, Braunholz DA. Lilford RJ, *et al.* Ethical issues in the design and conduct of cluster randomized controlled trials. *BMJ* 1999;318:1407–1409.
13. Martin DC, Diehr P, Perrin EB, Koepsell TD. The effect of matching on the power of randomized community intervention studies. *Statist Med* 1993;12:329–338.
14. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* 2003;327:785–789.
15. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. New York: John Wiley & Sons; 2012.
16. Siebenhofer A, Paulitsch MA, Pregartner G, Berghold A, Jeitler K, Muth C, Engler J. Cluster-randomized controlled trials evaluating complex interventions in general practices are mostly ineffective: a systematic review. *J Clin Epidemiol* 2018;94:85–96.
17. Campbell MK, Piaggio G, Elbourne DR, Altman DG; for the CONSORT Group. *Consort 2010 statement: extension to cluster randomised trials*. BMJ 2012;345:e5661.
18. Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. *BMJ* 2017;358:j3064.
19. Reich NG, Myers JA, Obeng D, Milstone AM, Perl TM. Empirical power and sample size calculations for cluster-randomized and cluster-randomized crossover studies. *PLoS One* 2012;7(4):e35564.
20. Caille A, Kerry S, Tavernier E, Leyrat C, Eldridge S, Giraudeau B. Timeline cluster: a graphical tool to identify risk of bias in cluster randomised trials. *BMJ* 2016;354:i4291.

https://doi.org/10.1017/ice.2019.48 Published online by Cambridge University Press